Cytometry Metadata in XML (BiOS 9711-30)

Robert C. Leif[1*], Stephanie H. Leif[1]

[1]XML_Med, a Division of Newport Instruments, 3345 Hopi Place, San Diego, CA, rleif@rleif.com

The CytometryML XML schemas and XML pages is available at www.cytometryml.org. You can read this presentation there at your leisure. You can also download the schemas and comment, suggest, improve, and/or use them.

What is MIFlowCyt?

- Presently, it is a file or a short appendix to a scientific article, such as one published in Cytometry.
- MIFlowCyt provides a concise description of the technology and findings present in the article.
- MIFlowCyt need **not** contain sufficient information to repeat the experiment, such as would be provided by a well written Materials and Method section of a paper and/or supplementary materials.
- The information contained in a MIFlowCyt document can be textural and/or structured.

Elements Are Required for the Real-World

```
<complexType name="CAS_Num_or_Other_Value_Type"
  mixed="true">
  <choice>
    <element name="CAS_Num" type="cas:CAS_Num_Type"/>
     <element name="ISAC_Num" type="cas:ISAC_Num_Type"/>
     <element name="Vendor_Num“
        type="strings:Bounded_128_Type"/>
      <element name="Other_Value“
        type="strings:Bounded_128_Type"/>
   </choice>
</complexType>
```

MIFlowCyt_Info

```
<complexType name="Miflowcyt_Info_Type" mixed="true">
<!—about 18 elements→
   <sequence>
      <element name="Experiment_Overview"
         type="exper_overview:Experiment_Overview_Type"/>
      <element name="Sample_Info“
         type="sample:Sample_Info_Type"/>
      <element name="Sample_Treatment_Description"
         type="protocol:Sample_Treatment_Description_Type"/>
```

MIFlowCyt_Info Continued

```
<choice>
 <element name="Flow_Series_and_Instance_Info"
      type="flow:Flow_Series_and_Instance_Info_Type"/>
 <element
      name="Microscope_Series_and_Instance_Info"
      type=
      "micro:Microscope_Series_and_Instance_Info_Type"/>
</choice> </sequence> </complexType>
```

MIFlowCyt.xsd Status

1. MIFlowCyt_Info is a significant part of MIFlowCyt, which has been coded in XSD1.1. and with elements present translates into 882 lines of XML. It should be able to be the basis of the materials and methods of a structured document.
2. When only the required elements were included, it validates with 432 lines of code, which should be suitable for the present synopsis.
3. MIFlowCyt.xml is based upon a tree of approximately 77 schemas.

- The next step in this project is to finish the Sample_Treatment_Description_Type

Problem: How does one verify that a MIFlowCyt section of a publication is complete?

- I believe that this presently is highly subjective.
- The reference example, Blimkie et al. "Identification of B cells through negative gating—an example of the MIFlowCyt standard applied", Cytometry Part A; 2010: 77A:546–551 Supplement is the exemplar. Does its text validate with the CytometryML schemas?
- Since the text is 9 pages of 12 point type, it is a considerable part of a Cytometry paper, which is duplicative.
- If there is a discrepancy, which is nominative the MIFlowCyt or the text of the paper?

What is to be done with the XML files generated with the CytometryML schemas?

- Since Flow Cytometry Standard, FCS, works on individual specially formatted files, you need special software to read your own data
- DICOM can work with a database and the applications that access that database. Because of FCS, ISAC members are used to moving the entire data set; rather, than using individual elements.

Opportunity Requires Instantiation of XHTML5

- **Opportunity**: Use a combination of XHTML and CSS tools to provide styles for the CytometryML code.
  - ❖Ubiquitous and widespread familiarity with HTML formatting elements (p, h1, h2, td, etc.)
- Formatting elements should be orthogonal to the data elements in XML applications including CytometryML and the DSTF XML.
- **Problem**: XHTML5 was not written in an XML schema language.

XML & XHTML5 presently do not interoperate

- **Solution**: Find a third party XML schema for XHTML 5, which is the one authored by Olivier Ishacian, Copyright (c) 2012 Pixware SARL.
- The elements of both schemas have to be oblivious of each other.
- The XHTML5 schema contributes the formatting elements.

Example

- Since MiflowCyte includes the reagents used, it is necessary to describe the fluorescent dyes used for a measurement. The ISAC DSTF has been discussing this.
- Obviously since there are many dyes, some sort of **tabular representation** should be included. One item of interest is the Chemical Abstracts Service (**CAS**) number, which is a **unique id**, such as 3301-79-9. As with all entries, the integrity of the database must be maintained by use of a string that matches a specified pattern

Namespace-based Validation Dispatching language (NVDL) Example

```
<?xml version="1.0" encoding="UTF-8"?>
<rules
xmlns="http://purl.oclc.org/dsdl/nvdl/ns/structure/1.
0"
xmlns:a="http://relaxng.org/ns/compatibility/annotati
ons/1.0"
     startMode="start">
      <mode name="start">
         <namespace
 ns="http://www.cytometryml.org/ACS/cas">
     <validate schema="cas.xsd" useMode="allAddress"/>
         </namespace>
     </mode>
```

HTML CAS Table Simulation

Common Name	Abbreviation	Excitation Max	Emission Max	Source URI
6-carboxyfluorescein	6-FAM	492	518	3301-79-9
6-JOE	6-JOE	520	550	82855-40-1

Microsoft Edge for Windows 10

Dye List 28 Jan. 2016.xml

Common Name	Abbreviation	Excitation Max	Emission Max	URI	CAS NUM
6-carboxyfluorescein	6-FAM	492	518	http://www.cheap.com/	3301-79-9
Joe	6-Joe	520	550	http://www.cheap.com/	82855-40-1

FireFox

Dye List 28 Jan. 2016.xml

Common Name	Abbreviation	Excitation Max	Emission Max	URI	CAS NUM
6-carboxyfluorescein	6-FAM	492	518	http://www.cheap.com/	3301-79-9
Joe	6-Joe	520	550	http://www.cheap.com/	82855-40-1

Conclusions

1. It is possible to create XHTML5 files that include both formatting and strongly typed data elements.
2. Since XHTML files can be included in the form of EPUB files and EPUB files can store images and other binary data, it should be possible to use EPUBs as ISAC Archival Cytometry Standard files.
3. The processing of XML by CSS stylesheets needs to be improved.
4. MIFlowCyt can serve as a table of contents, ToC, with hypertext linkages to the detailed descriptions, which will be based on web pages generated from the CytometryML schemas. This material can be contained in the paper or supplementary material.
5. Interested Parties: reviewers, editors, government agencies (FDA) can automatically check for completeness in answering their requirements by validating the ToC against its schema(s).

- **Reference**

- **NVDL Tutorial,** http://jnvdl.sourceforge.net/tutorial.html#d4e142

Acknowledgments

This work was sponsored by Newport Instruments internal research funds. One of us, R.C.L. wishes to thank Ryan R. Brinkman, Josef Spidlen and the other members of the ISAC DSTF for many enlightening and pleasant discussions and particularly Kim Blenman who performed the work on obtaining and preparing CAS numbers. Radu Coravu of Syncro Soft SRL, the manufacturer of oXygen (www.oxygenxml.com), suggested and facilitated the use of a meta-schema language control language, NVDL. Olivier Ishacian and his company, Pixware SARL, for making the XHTML5 XSD schema available to the authors and the rest of the public.